PCT/DK 2004/000468

10/562021

Kongeriget Danmark

Patent application No.: PA 2003 00994

Date of filing: 01 July 2003

Applicant: (Name and address)
NOVOZYMES A/S
Krogshøjvej 36
DK-2880 Bagsværd
Denmark

Title: Cgtase variants

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

16 July 2004

Susanne Morsing

BEST AVAILABLE COPY

PATENT- OG VAREMÆRKESTYRELSEN

CGTASE VARIANTS

FIELD OF THE INVENTION

The present invention relates to the construction of variants of cyclodextrin glucanotransferases (CGTases), in particular variants having the ability to form linear oligosaccharides.

BACKGROUND OF THE INVENTION

Pdb files 1CDG, 1PAM, 1CYG and 1CIU (available at www.rcsb.org) show the amino acid sequences and three-dimensional structures of several cyclodextrin glucanotransferases (CGTases). WO 9943794 shows the amino acid sequence and three-dimensional structure of a maltogenic alpha-amylase from *Bacillus stearothermophilus*, known as Novamyl ®.

Variants of a cyclodextrin glucanotransferase (CGTase) with the ability to form linear oligosaccharides are disclosed in WO 9943793 and in R.J. Leemhuis: "What makes cyclodextrin glycosyltransferase a transglycosylase", University Library Groningen, 2003.

L. Beier et al., Protein Engineering, vol 13, no. 7, pp. 509-513, 2000 is titled "Conversion of the maltogenic α-amylase Novamyl into a CGTase".

SUMMARY OF THE INVENTION

The inventors have developed a method of modifying the amino acid sequence of a CGTase to obtain variants. The variants may form linear oligosaccharides as an initial product by starch hydrolysis and a reduced amount of cyclodextrin and may be useful for anti-staling in baked products.

Accordingly, the invention provides a method of constructing CGTase variants based on a comparison of three-dimensional (3D) structures of the CGTase and a maltogenic alpha-amylase. One or both models includes a substrate. The invention also provides novel CGTase variants.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1 shows the results of a comparison of the 3D structures 1a47 for a CGTase (SEQ ID NO: 2) and 1qho for the maltogenic alpha-amylase Novamyl (SEQ ID NO: 1). Details are described in Example 1.

DETAILED DESCRIPTION OF THE INVENTION

CGTase

The method of the invention uses an amino acid sequence of a CGTase and a three-dimensional model for the CGTase. The model may include a substrate.

The CGTase may have a three-dimensional structure found under the indicated identifier in the Protein Data Bank (www.rcsb.org): *B. circulans* (1CDG), alkalophilic *Bacillus* (1PAM), *B. stearothermophilus* (1CYG) or *Thermoanaerobacterium thermosulfurigenes* (1CIU, 1A47). 3D structures for other CGTases may be constructed as described in Example 1 of WO 9623874.

The CGTase may particularly have a sequence as found under the following accession numbers in the GeneSeqP database for CGTase from the indicated microorganism:

1. aab71493.gcg *B. agaradherens*
2. aau76326.gcg *Bacillus agaradhaerans*
3. cdg2_paema.gcg *Paenibacillus macerans* (*Bacillus macerans*).
4. cdg1_paema.gcg *Paenibacillus macerans* (*Bacillus macerans*).
5. cdgt_thetu.gcg *Thermoanaerobacter thermosulfurogenes* (*Clostridium thermosulfurogenes*)
6. aaw06772.gcg *Thermoanaerobacter thermosulphurigenes* sp. ATCC 53627
7. cdgt_bacci.gcg *Bacillus circulans*
8. cdgt_bacli.gcg *Bacillus sp.* (strain 38-2)
9. cdgt_bacs3.gcg *Bacillus sp.* (strain 38-2)
10. cdgt_bacs0.gcg *Bacillus sp.* (strain 1011)
11. cdgu_bacci.gcg *Bacillus circulans*
12. cdgt_bacsp.gcg *Bacillus sp.* (strain 17-1)
13. cdgt_bacst.gcg *Bacillus stearothermophilus*
14. cdgt_bacoh.gcg *Bacillus ohbensis*
15. cdgt_bacs2.gcg *Bacillus sp.* (strain 1-1)
16. cdgt_klepn.gcg *Klebsiella pneumoniae*

To develop variants of a CGTase without a known 3D structure, the sequence may be aligned with a CGTase having a known 3D structure. The sequence alignment may be done by conventional methods, e.g. by use the software GAP from UWGCG Version 8.

Maltogenic alpha-amylase

The method also uses an amino acid sequence of a maltogenic alpha-amylase (EC 3.2.1.133) and a three-dimensional model of the maltogenic alpha-amylase. The model may include a substrate. The maltogenic alpha-amylase may have the amino acid sequence have the amino acid sequence shown in SEQ ID NO: 1 (in the following referred to as Novamyl). A 3D model for Novamyl with a substrate is described in US 6162628 and is found in the Protein Data Bank with the identifier 1QHO. Alternatively, the maltogenic alpha-amylase may be a No-

vamyl variant described in US 6162628. A 3D structure of such a variant may be developed from the Novamyl structure by known methods, e.g. as described in T.L. Blundell et al., Nature, vol. 326, p. 347 ff (26 March 1987); J. Greer, Proteins: Structure, Function and Genetics, 7:317-334 (1990); or Example 1 of WO 9623874.

Superimposition of 3D models

The two 3D models may be superimposed by aligning the amino acid residues of each catalytic triad. This may be done by methods known in the art based on the deviations of the three pairs of C-alpha atoms, e.g. by minimizing the sum of squares of the three deviations or by aligning so as to keep each deviation below 0.8 Å, e.g. below 0.6 Å, below 0.4 Å,, below 0.3 Å or below 0.2 Å.

Alternatively, the superimposition may be based on the deviations of all corresponding pairs of amino acid residues as shown in the alignment in Figs. 4-5 of WO 9943793 and bringing the sum of square of all deviations to a minimum.

Selection of amino acid sequences

In the superimposed 3D models, amino acid residues in the CGTase sequence are selected by two criteria: Firstly, CGTase residues < 10 Å from a substrate (having a C-alpha atom located < 10 Å from an atom of a substrate) are selected. Secondly, CGTase residues > 0.8 Å from any maltogenic alpha-amylase residue (having a C-alpha atom > 0.8 Å from the C-alpha atom of any maltogenic alpha-amylase residue) are selected.

Modifications of CGTase amino acid sequence

One or more of the following modifications are made to the CGTase sequence:

Deletion or substitution

A CGTase residue < 10 Å from a substrate and > 0.8 Å from any residue in the maltogenic alpha-amylase sequence may be deleted or may be substituted with a different residue.

The substitution may be made with the same amino acid residue as found at a corresponding position in the maltogenic alpha-amylase sequence or with a residue of the same type. The type indicates a positively charged, negatively charged, hydrophilic or hydrophobic residue, understood as follows (Tyr may be hydrophilic or hydrophobic):

Hydrophobic amino acids: Ala, Val, Leu, Ile, Pro, Phe, Trp, Gly, Met, Tyr

Hydrophilic amino acids: Thr, Ser, Gln, Asn, Tyr, Cys

Positively charged amino acids: Lys, Arg, His

Negatively charged amino acids: Glu, Asp

The CGTase residue may be substituted with a larger or smaller residue depending on whether a larger or smaller residue is found at a corresponding position in the maltogenic alpha-amylase sequence. In this connection, the residues are ranked as follows from smallest to largest: (an equal sign indicates residues with sizes that are practically indistinguishable):

G < A=S=C < V=T < P < L=I=N=D=M < E=Q < K < H < R < F < Y < W

Also, a stretch (a "loop") of consecutive CGTase residues may be selected if each of the residues is > 0.8 Å from any residue in the maltogenic alpha-amylase sequence and some of the CGTase residues is <10 Å from a substrate. Such a stretch of CGTase residues may be deleted or substituted with different amino acid residues. The substitution may be made with the residues found at the corresponding location in the maltogenic alpha-amylase sequence, with residues of the same type, or with an equal number of residues or one or two more or fewer residues than found in the maltogenic alpha-amylase sequence.

Insertion

One or more amino acid residues may be inserted at a position in the CGTase sequence corresponding to one or more residues in the maltogenic alpha-amylase sequence which < 10 Å from a substrate and which are > 0.8 Å from any CGTase residue. The insertion may be made with the same residue or with an amino acid residue of the same type as the amino acid residue in the maltogenic alpha-amylase sequence. The type indicates a positively charged, negatively charged, hydrophilic or hydrophobic residue, as above.

Where the maltogenic alpha-amylase sequence contains a stretch (a peptide loop) of residues < 10 Å from a substrate and > 0.8 Å from any CGTase residue, the insertion at the corresponding position in the CGTase sequence may consist of an equal number of residues, or the insertion may have one or two fewer or more residues. Thus, in the case of a stretch of 5 such residues in the maltogenic alpha-amylase sequence, the insertion may be made with 1-7 residues, e.g. 1, 2, 3, 4, 5, 6 or 7 residues. Each inserted residue may be the same as one of the maltogenic alpha-amylase residues or of the same type.

Optional further modifications of the CGTase sequence

Optionally, the CGTase sequence may be further modified by substituting one or more residues which is matched with a residue in the maltogenic alpha-amylase sequence.

The substitution may be made with an amino acid residue of the same type (in particular with the same residue) as the matching residue in the maltogenic alpha-amylase sequence.

Depending on whether the matching residue in the maltogenic alpha-amylase sequence is smaller or larger than the residue in the CGTase sequence, the substitution may be made with a smaller or larger residue (using the ranking shown above).

Production of CGTase variants

A polypeptide having the resulting amino acid sequence may be produced by conventional methods, generally involving producing DNA with a sequence encoding the polypeptide together with control sequences, transforming a suitable host organism with the DNA , cultivating the transformed organism at suitable conditions for expressing and optionally secreting the polypeptide, and optionally recovering the expressed polypeptide.

DNA encoding any of the above CGTase variants may be prepared, e.g. by point-specific mutation of DNA encoding the parent CGTase. This may be followed by transformation of a suitable host organism with the DNA, and cultivation of the transformed host organism under suitable conditions to express the encoded polypeptide (CGTase variant). This may be done by known methods.

Optional screening of CGTase variants

Optionally, one or more expressed polypeptides may be tested for one or more useful enzymatic activities. This may include testing for the ability to hydrolyze starch or a starch derivative by a conventional method, e.g. a plate assay, use of Phadebas tablets or DSC on amylopectin. Further, the initial product from starch hydrolysis may be analyzed and a polypeptide producing an increased ratio of linear oligosaccharides to cyclodextrins may be selected. Also, the polypeptide may be tested by adding it to a dough, baking it and testing the firmness of the baked product during storage; a polypeptide with anti-staling effect may be selected as described in WO 9104669 or US 6162628. Finally, the polypeptide may be tested for thermostability, and a more thermostable one may be preferred.

Optional gene recombination

Optionally, DNA encoding a plurality of the above CGTase variants may be prepared and recombined, followed by transformation of a suitable host organism with the recombined DNA, and cultivation of the transformed host organism under suitable conditions to express the encoded polypeptides (CGTase variants). The gene recombination may be done by known methods.

CGTase variants

Particularly, the CGTase may be modified by substitution, insertion or deletion of an amino acid at a position corresponding to amino acid 85-95, 152, 184, 260-269, 285, 288, 314 of the amino acid sequence shown in SEQ ID NO: 2 or 3. The modification may comprise substitution or insertion of an amino acid residue with an amino acid residue of a corresponding position in the amino acid sequence of Novamyl (SEQ ID NO: 1) or a deletion of an amino acid

residue in the region which is not present at the corresponding position in the Novamyl sequence.

More particularly, the modification may comprise substitution of amino acids corresponding to amino acids 85-95, 260-268 or 260-269 of SEQ ID NO: 2 or 3 with TLAGTDN, YGDDPGTANHL or YGDDPGTANHLE, respectively.

Some particular examples with the *Thermoanaerobacter* CGTase (SEQ ID NO: 3) as an example are Y152F, F184W, R285D, Q288T, D314E. Corresponding substitutions may be made in other CGTases.

Also, one or more additional modifications may be made, each being an amino acid substitution, insertion or deletion. In particular, such modification may be made in the regions corresponding to amino acids 40-43, 78-85, 136-139, 173-180, 189-195 or 258-268 of SEQ ID NO: 1. In particular, the modification may be an insertion of or a substitution with an amino acid present at the corresponding position of Novamyl, or a deletion of an amino acid not present at the corresponding position of Novamyl. Thus, taking the *Thermoanaerobacter* CGTase (SEQ ID NO: 3) as an example, one or more of the following changes may be made to introduce a loop modeled on Novamyl:

- A85-S95 of SEQ ID NO: 3 is replaced by T80-N86 of SEQ ID NO: 1,
- N194-L198 of SEQ ID NO: 3 is replaced by N187-L196 of SEQ ID NO: 1,
- Y260-P268 of SEQ ID NO: 3 is replaced by Y258-L268 of SEQ ID NO: 1, or
- Y260-N269 of SEQ ID NO: 3 is replaced by Y258-E269 of SEQ ID NO: 1.

The following are particular examples of variants based on the *Thermoanaerobacter* CGTase (SEQ ID NO: 3):

Variant 1: Loop A85-S95 to Novamyl loop T80-N86, Loop N194-L198 to Novamyl Loop N187-L196, and Y152F

Variant 2: As Variant 1 with addition of F184W, R285D, Q288T, D314E, and Loop Y260-P268 to Novamyl Loop Y258-L268.

Variant 3: Loop A85-S95 to Novamyl loop T80-N86, Loop Y260-P268 to Novamyl loop Y258-L268, Y152F, G257D, R285D, Q288T, D314E.

EXAMPLES

Example 1: Construction of CGTase residues based on 3D structures

Two 3D structures with substrates were used: 1A47 for a CGTase (SEQ ID NO: 2) and 1 QHO for a maltogenic alpha-amylase (Novamyl, SEQ ID NO: 1), wherein the substrates are indicated as GTE, GLC, CYL and GLD for 1a47 and as ABD for 1 qho. The two structures were superimposed by minimizing the sum of squares for deviations at the three C-alpha atoms at the catalytic triad: D230, E258 and D329 for 1A47, and D228, E256 and D329 for Novamyl. The

superimposed structures were analyzed, and the result is shown in Fig. 1 with the Novamyl sequence at the top and the CGTase sequence below.

The following CGTase residues were found to have a C-alpha atom < 10 Å from an atom of either substrate: 19, 21, 24, 46-47, 75, 77-78, 82-83, 85-103, 106, 136-145, 152-153, 182-187, 190-191, 193-200, 228-235, 257-267, 270, 282-289, 291-292, 296, 298, 324, 327-331, 359, 369-375. They are indicated by the first underlining in Fig. 1.

Two stretches ("loops") of consecutive residues were identified where some residues have the C-alpha atom < 10 Å from an atom of either substrate and > 0.8 Å from the C-alpha atom of any Novamyl residue. Including prefix and postfix, the two stretches are at residues 85-96 and 193-200 of the CGTase.

The following CGTase residues were found to be included in either of the above subsets (<10 Å from a substrate or in a loop) and to have a C-alpha atom > 0.8 Å from the C-alpha atom of any Novamyl residue: 75, 77, 78, 85-94, 140, 144-145, 152, 182-187, 193-197, 235, 262-266, 286-289, 292, 296, 298, 369-370. They are indicated by the second underlining in Fig. 1.

Variants were constructed by selecting residues in the CGTase of SEQ ID NO: 2 from residues with the second underlining in Fig. 1 and identifying the corresponding residues in the CGTase of SEQ ID NO: 3 from an alignment of the two CGTase sequences. As a result of the high degree of identity, the residues have the same numbers in the two sequences. The selected residues in SEQ ID NO: 3 were substituted as indicated below.

Variant 1 was created from SEQ ID NO: 3 as follows: CGTase residues A85-S95 were substituted with Novamyl residues T80-N86. CGTase residues N194-L198 were substituted with Novamyl residues N187-L196. Further, substitution Y152F was made to the CGTase sequence.

Variant 2 was created as Variant 1 with the following additional substitutions in SEQ ID NO: 3: CGTase residues Y260-P268 were substituted with Novamyl residues Y260-P268. Further substitutions F184W, R285D, Q288T, and D314E were made to the CGTase sequence.

Variant 3 was created from SEQ ID NO: 3 as follows: CGTase residues A85-S95 were substituted with Novamyl residues T80-N86. CGTase residues L261-P268 were substituted with Novamyl residues D261-L268. The following further substitutions were made in the CGTase sequence: Y152F, G257D, R285D, Q288T and D314E.

CLAIMS

1. A method of producing a variant polypeptide, which method comprises:

a) providing an amino acid sequence and a three-dimensional model for a cyclodextrin glucanotransferase (CGTase) and for an amino acid sequence for a maltogenic alpha-amylase wherein one or both models includes a substrate,

b) superimposing the two three-dimensional models,

c) and modifying the amino acid sequence of the CGTase wherein the modification comprises:

i) deleting an amino acid residue in the CGTase sequence which has a C-alpha atom < 10 Å from an atom of a substrate and > 0.8 Å from the C-alpha atom of any amino acid residue in the maltogenic alpha-amylase sequence,

ii) substituting an amino acid residue in the CGTase sequence which has a C-alpha atom < 10 Å from an atom of a substrate and > 0.8 Å from the C-alpha atom of any amino acid residue in the maltogenic alpha-amylase sequence with a different amino acid residue, or

iii) deleting or substituting a stretch of consecutive CGTase residues wherein each residue is > 0.8 Å from any residue in the maltogenic alpha-amylase sequence and comprising at least one CGTase residue <10 Å from a substrate,

iv) inserting an amino acid residue at a position in the CGTase sequence corresponding to a maltogenic alpha-amylase sequence which has a C-alpha atom < 10 Å from an atom of a substrate and > 0.8 Å from the C-alpha atom of any CGTase residue, and

d) producing the polypeptide having the resulting amino acid sequence.

2. The method of claim 1 wherein the substitution is made with an amino acid residue of the same type as an unmatched amino acid residue at a corresponding position in the maltogenic alpha-amylase sequence, wherein the type is positively charged, negatively charged, hydrophilic or hydrophobic.

3. The method of claim 1 wherein the insertion is made with an amino acid residue of the same type as an unmatched amino acid residue at a corresponding position in the maltogenic alpha-amylase sequence, wherein the type is positively charged, negatively charged, hydrophilic or hydrophobic.

4. The method of any preceding claim wherein the modification of the amino acid sequence further comprises substitution of a matched amino acid residue in the CGTase sequence which has a C-alpha atom located less than 10 Å from an atom of a substrate with a different amino acid residue.

5. The method of the preceding claim wherein the substitution of the matched amino acid residue is made with an amino acid residue of the same type as the matching amino acid residue of the maltogenic alpha-amylase sequence, wherein the type is positively charged, negatively charged, hydrophilic or hydrophobic.

6. The method of any preceding claim which further comprises preparing the variant polypeptide, letting it act on starch, and selecting a variant polypeptide having the ability to form linear oligosaccharide as an initial product.

7. A polypeptide which:

a) has an amino acid sequence having at least 70% identity to a parent cyclodextrin glucanotransferase (CGTase);

b) comprises insertion of an amino acid compared to the parent CGTase in a region corresponding to amino acids 194-198 of SEQ ID NO: 3,

c) comprises an amino acid modification compared to the parent CGTase which is substitution, insertion or deletion of an amino acid at a position corresponding to amino acid 85-95, 152, 184, 260-269, 285, 288, 314 of the amino acid sequence shown in SEQ ID NO: 3, and

d) has the ability to form linear oligosaccharides as an initial product when acting on starch.

8. The polypeptide of claim 7 comprising insertion of 1-7 amino acids, particularly 5 amino acids, more particularly insertion of DPAGF, most particularly between amino acids corresponding to 196 and 197 of SEQ ID NO: 3.

9. The polypeptide of claim 7 or 8, further comprising substitution of an amino acid corresponding to any of amino acids 194-198 of SEQ ID NO: 3, particularly a substitution corresponding to L195F, F196T or D197S in SEQ ID NO: 3.

10. The polypeptide of any preceding claim, comprising a substitution or insertion of an amino acid residue with an amino acid residue of a corresponding position in the amino acid sequence shown in SEQ ID NO: 1 or a deletion of an amino acid residue in the region which is not present at the corresponding position in the amino acid sequence shown in SEQ ID NO: 1.

11. The polypeptide of any preceding claim, comprising substitution of amino acids corresponding to amino acids 85-95, 260-268 or 260-269 of SEQ ID NO: 3 with TLAGTDN, YGDDPGTANHL or YGDDPGTANHLE, respectively.

12. The polypeptide of any preceding claim comprising a substitution corresponding to Y152F, F184W, R285W, Q288T, D314E in SEQ ID NO: 3.

13. A process for preparing a baked product which comprises adding the polypeptide of any preceding claim, or a polypeptide produced by the method of any of claims 1-6 to a dough and baking the dough to prepare the baked product, wherein the polypeptide is added in an amount which is effective to retard the staling of the baked product.

10340 Fig. 1

```
1         10        20        30        40        50        60        70
------SSSASVKGDVIYQIIIDRFYDGDTTNNNPAKSYGLYDPTKSKWKMYWGGDLEGVRQKL--PYLK
ASDTAVSNVVNYSTDVIYQIVTDRFVDGNTSNNPT---GDLYDPTHTSLKKYFGGDWQGIINKINDGYLT 67
                    - -  -                          ---

                  ************
PVLDNLDTLAGT----DNTGYHGYWTRDFKQIEEHFGNWTTFDTLVNDAHQNGIKVIVDFVPNHSTPFKA
QPVENIYAVLPDSTFGGSTSYHGYWARDFKRTNPYFGSFTDFQNLINTAHAHNIKVIIDFAPNHTSPASE 137
        - --   --   -  - -- ----------   -                              --
        - --        - - --  -

                                                          **************
NDSTFAEGGAQLGVTTIWLSLYNNGTYMGNYFDDATKGYFHHNGDISNWDDRYEAQWKNFTDPAGFSLAD
TDPTYAENGRGMGVTAIWISLYDNGTLLGGYTNDT-NGYFHHYGGTD-FSSYEDGIYRNLF-----DLAD 200
--------      --                                -- ---- --- ----      ----
  -   --                                          -- ---- -   ----      -

LSQENGTIAQYLTDAAVQLVAHGADGLRIDAVKHFNSGFSKSLADKLYQKKDIFLVGEWYGDD-PGTANH
LNQQNSTIDSYLKSAIKVWLDMGIDGIRLDAVKHMPFGWQKNFMDSILSYRPVFTFGEWFLG-TNEI--D 267
                                 --------                        ------ ---- -
                                        -                            - ----

LEKVRYANNSGVNVLDFDLNTVIRNVFGTFTQTMYDLNNMVNQTGNEYKYKENLITFIDNHDMSRFLSVN
VNNTYFANESGMSLLDFRFSQKVRQVFRDNTDTMYGLDSMIQSTASDYNFINDMVTFIDNHDMDRFYN-G 336
  -           - -- --- --  -                                  -----
                    ---  -   -

SNKANLHQALAFILTSRGTPSIYYGTEQYMAGGNDPYNRGMMPAFDTTTTAFKEVSTLAGLRRNNAAIQY
GSTRPVEQALAFTLTSRGVPAIYYGTEQYMTGNGDPYNRAMMTSFNTSTTAYNVIKKLAPLRKSNPAIAY 406
                          -         -------
                                    --

GTTTQRWINNDVYIYERKFFNDVVLVAINRNTQSSYSISGLQTALPNGSYADYLSGLLGGNGISVS-NGS
GTTQQRWINNDVYIYERKFGNNVALVAINRNLSTSYNITGLYTALPAGTYTDVLGGLLNGNSISVASDGS 476

VASFTLAPGAVSVWQYST-SASAPQIGSVAPNMGIPGNVVTIDGKGFGTTQGTVTFGGVTATVKSWTSNR
VTPFTLSAGEVAVWQYVSSSN-SPLIGHVGPTMTKAGQTITIDGRGFGTTSGQVLFGSTAGTIVSWDDTE 545

IEVYVPNMAAGLTDVKVTA-GGVSSNLYS-YNILSGTQTSVVFTVKSAPPTNLGDKIYLTGNIPELGNWS
VKVKVPSVTPGKYNISLKTSSGATSNTYNNINILTGNQICVRFVVNNASTVY-GENVYLTGNVAELGNWD 614

TDTSGAVNNAQGPLLAP---NYPDWFYVFSVPAGKTIQFKFFIKRADGT-IQWENGSNHVATTPTGATGN
TS------KAIGPMFNQVVYQYPTWYYDVSVPAGTTIQFKFIKKN--GNTITWEGGSNHTYTVPSSSTGT 676

ITVTWQN
VIVNWQQ 683
```

Fig. 1

SEQUENCE LISTING

<110> Novozymes A/S

<120> CGTASE VARIANTS

<130> 10340-000

<160> 3

<170> PatentIn version 3.2

<210> 1
<211> 686
<212> PRT
<213> Bacillus stearothermophilus

<400> 1

Ser Ser Ser Ala Ser Val Lys Gly Asp Val Ile Tyr Gln Ile Ile Ile
1 5 10 15

Asp Arg Phe Tyr Asp Gly Asp Thr Thr Asn Asn Asn Pro Ala Lys Ser
20 25 30

Tyr Gly Leu Tyr Asp Pro Thr Lys Ser Lys Trp Lys Met Tyr Trp Gly
35 40 45

Gly Asp Leu Glu Gly Val Arg Gln Lys Leu Pro Tyr Leu Lys Gln Leu
50 55 60

Gly Val Thr Thr Ile Trp Leu Ser Pro Val Leu Asp Asn Leu Asp Thr
65 70 75 80

Leu Ala Gly Thr Asp Asn Thr Gly Tyr His Gly Tyr Trp Thr Arg Asp
85 90 95

Phe Lys Gln Ile Glu Glu His Phe Gly Asn Trp Thr Thr Phe Asp Thr
100 105 110

Leu Val Asn Asp Ala His Gln Asn Gly Ile Lys Val Ile Val Asp Phe
115 120 125

Val Pro Asn His Ser Thr Pro Phe Lys Ala Asn Asp Ser Thr Phe Ala
130 135 140

Glu Gly Gly Ala Leu Tyr Asn Asn Gly Thr Tyr Met Gly Asn Tyr Phe
145 150 155 160

Asp Asp Ala Thr Lys Gly Tyr Phe His His Asn Gly Asp Ile Ser Asn
165 170 175

Trp Asp Asp Arg Tyr Glu Ala Gln Trp Lys Asn Phe Thr Asp Pro Ala
180 185 190

Gly Phe Ser Leu Ala Asp Leu Ser Gln Glu Asn Gly Thr Ile Ala Gln

195 200 205

Tyr Leu Thr Asp Ala Ala Val Gln Leu Val Ala His Gly Ala Asp Gly
210 215 220

Leu Arg Ile Asp Ala Val Lys His Phe Asn Ser Gly Phe Ser Lys Ser
225 230 235 240

Leu Ala Asp Lys Leu Tyr Gln Lys Lys Asp Ile Phe Leu Val Gly Glu
245 250 255

Trp Tyr Gly Asp Asp Pro Gly Thr Ala Asn His Leu Glu Lys Val Arg
260 265 270

Tyr Ala Asn Asn Ser Gly Val Asn Val Leu Asp Phe Asp Leu Asn Thr
275 280 285

Val Ile Arg Asn Val Phe Gly Thr Phe Thr Gln Thr Met Tyr Asp Leu
290 295 300

Asn Asn Met Val Asn Gln Thr Gly Asn Glu Tyr Lys Tyr Lys Glu Asn
305 310 315 320

Leu Ile Thr Phe Ile Asp Asn His Asp Met Ser Arg Phe Leu Ser Val
325 330 335

Asn Ser Asn Lys Ala Asn Leu His Gln Ala Leu Ala Phe Ile Leu Thr
340 345 350

Ser Arg Gly Thr Pro Ser Ile Tyr Tyr Gly Thr Glu Gln Tyr Met Ala
355 360 365

Gly Gly Asn Asp Pro Tyr Asn Arg Gly Met Met Pro Ala Phe Asp Thr
370 375 380

Thr Thr Thr Ala Phe Lys Glu Val Ser Thr Leu Ala Gly Leu Arg Arg
385 390 395 400

Asn Asn Ala Ala Ile Gln Tyr Gly Thr Thr Thr Gln Arg Trp Ile Asn
405 410 415

Asn Asp Val Tyr Ile Tyr Glu Arg Lys Phe Phe Asn Asp Val Val Leu
420 425 430

Val Ala Ile Asn Arg Asn Thr Gln Ser Ser Tyr Ser Ile Ser Gly Leu
435 440 445

Gln Thr Ala Leu Pro Asn Gly Ser Tyr Ala Asp Tyr Leu Ser Gly Leu
450 455 460

Leu Gly Gly Asn Gly Ile Ser Val Ser Asn Gly Ser Val Ala Ser Phe

465 470 475 480

Thr Leu Ala Pro Gly Ala Val Ser Val Trp Gln Tyr Ser Thr Ser Ala
485 490 495

Ser Ala Pro Gln Ile Gly Ser Val Ala Pro Asn Met Gly Ile Pro Gly
500 505 510

Asn Val Val Thr Ile Asp Gly Lys Gly Phe Gly Thr Thr Gln Gly Thr
515 520 525

Val Thr Phe Gly Gly Val Thr Ala Thr Val Lys Ser Trp Thr Ser Asn
530 535 540

Arg Ile Glu Val Tyr Val Pro Asn Met Ala Ala Gly Leu Thr Asp Val
545 550 555 560

Lys Val Thr Ala Gly Gly Val Ser Ser Asn Leu Tyr Ser Tyr Asn Ile
565 570 575

Leu Ser Gly Thr Gln Thr Ser Val Val Phe Thr Val Lys Ser Ala Pro
580 585 590

Pro Thr Asn Leu Gly Asp Lys Ile Tyr Leu Thr Gly Asn Ile Pro Glu
595 600 605

Leu Gly Asn Trp Ser Thr Asp Thr Ser Gly Ala Val Asn Asn Ala Gln
610 615 620

Gly Pro Leu Leu Ala Pro Asn Tyr Pro Asp Trp Phe Tyr Val Phe Ser
625 630 635 640

Val Pro Ala Gly Lys Thr Ile Gln Phe Lys Phe Phe Ile Lys Arg Ala
645 650 655

Asp Gly Thr Ile Gln Trp Glu Asn Gly Ser Asn His Val Ala Thr Thr
660 665 670

Pro Thr Gly Ala Thr Gly Asn Ile Thr Val Thr Trp Gln Asn
675 680 685

<210> 2
<211> 683
<212> PRT
<213> Thermoanaerobacterium thermosulfurigenes

<400> 2

Ala Ser Asp Thr Ala Val Ser Asn Val Val Asn Tyr Ser Thr Asp Val
1 5 10 15

Ile Tyr Gln Ile Val Thr Asp Arg Phe Val Asp Gly Asn Thr Ser Asn
20 25 30

```
Asn Pro Thr Gly Asp Leu Tyr Asp Pro Thr His Thr Ser Leu Lys Lys
        35                  40                  45

Tyr Phe Gly Gly Asp Trp Gln Gly Ile Ile Asn Lys Ile Asn Asp Gly
    50                  55                  60

Tyr Leu Thr Gln Pro Val Glu Asn Ile Tyr Ala Val Leu Pro Asp Ser
65                  70                  75                  80

Thr Phe Gly Gly Ser Thr Ser Tyr His Gly Tyr Trp Ala Arg Asp Phe
                85                  90                  95

Lys Arg Thr Asn Pro Tyr Phe Gly Ser Phe Thr Asp Phe Gln Asn Leu
            100                 105                 110

Ile Asn Thr Ala His Ala His Asn Ile Lys Val Ile Ile Asp Phe Ala
        115                 120                 125

Pro Asn His Thr Ser Pro Ala Ser Glu Thr Asp Pro Thr Tyr Ala Glu
    130                 135                 140

Asn Gly Arg Gly Met Gly Val Thr Ala Ile Trp Ile Ser Leu Tyr Asp
145                 150                 155                 160

Asn Gly Thr Leu Leu Gly Gly Tyr Thr Asn Asp Thr Asn Gly Tyr Phe
                165                 170                 175

His His Tyr Gly Gly Thr Asp Phe Ser Ser Tyr Glu Asp Gly Ile Tyr
            180                 185                 190

Arg Asn Leu Phe Asp Leu Ala Asp Leu Asn Gln Gln Asn Ser Thr Ile
        195                 200                 205

Asp Ser Tyr Leu Lys Ser Ala Ile Lys Val Trp Leu Asp Met Gly Ile
    210                 215                 220

Asp Gly Ile Arg Leu Asp Ala Val Lys His Met Pro Phe Gly Trp Gln
225                 230                 235                 240

Lys Asn Phe Met Asp Ser Ile Leu Ser Tyr Arg Pro Val Phe Thr Phe
                245                 250                 255

Gly Glu Trp Phe Leu Gly Thr Asn Glu Ile Asp Val Asn Asn Thr Tyr
            260                 265                 270

Phe Ala Asn Glu Ser Gly Met Ser Leu Leu Asp Phe Arg Phe Ser Gln
        275                 280                 285

Lys Val Arg Gln Val Phe Arg Asp Asn Thr Asp Thr Met Tyr Gly Leu
    290                 295                 300
```

```
Asp Ser Met Ile Gln Ser Thr Ala Ser Asp Tyr Asn Phe Ile Asn Asp
305                 310                 315                 320

Met Val Thr Phe Ile Asp Asn His Asp Met Asp Arg Phe Tyr Asn Gly
                325                 330                 335

Gly Ser Thr Arg Pro Val Glu Gln Ala Leu Ala Phe Thr Leu Thr Ser
            340                 345                 350

Arg Gly Val Pro Ala Ile Tyr Tyr Gly Thr Glu Gln Tyr Met Thr Gly
        355                 360                 365

Asn Gly Asp Pro Tyr Asn Arg Ala Met Met Thr Ser Phe Asn Thr Ser
    370                 375                 380

Thr Thr Ala Tyr Asn Val Ile Lys Lys Leu Ala Pro Leu Arg Lys Ser
385                 390                 395                 400

Asn Pro Ala Ile Ala Tyr Gly Thr Thr Gln Gln Arg Trp Ile Asn Asn
                405                 410                 415

Asp Val Tyr Ile Tyr Glu Arg Lys Phe Gly Asn Asn Val Ala Leu Val
            420                 425                 430

Ala Ile Asn Arg Asn Leu Ser Thr Ser Tyr Asn Ile Thr Gly Leu Tyr
        435                 440                 445

Thr Ala Leu Pro Ala Gly Thr Tyr Thr Asp Val Leu Gly Gly Leu Leu
    450                 455                 460

Asn Gly Asn Ser Ile Ser Val Ala Ser Asp Gly Ser Val Thr Pro Phe
465                 470                 475                 480

Thr Leu Ser Ala Gly Glu Val Ala Val Trp Gln Tyr Val Ser Ser Ser
                485                 490                 495

Asn Ser Pro Leu Ile Gly His Val Gly Pro Thr Met Thr Lys Ala Gly
            500                 505                 510

Gln Thr Ile Thr Ile Asp Gly Arg Gly Phe Gly Thr Thr Ser Gly Gln
        515                 520                 525

Val Leu Phe Gly Ser Thr Ala Gly Thr Ile Val Ser Trp Asp Asp Thr
    530                 535                 540

Glu Val Lys Val Lys Val Pro Ser Val Thr Pro Gly Lys Tyr Asn Ile
545                 550                 555                 560

Ser Leu Lys Thr Ser Ser Gly Ala Thr Ser Asn Thr Tyr Asn Asn Ile
                565                 570                 575
```

Asn Ile Leu Thr Gly Asn Gln Ile Cys Val Arg Phe Val Val Asn Asn
580 585 590

Ala Ser Thr Val Tyr Gly Glu Asn Val Tyr Leu Thr Gly Asn Val Ala
595 600 605

Glu Leu Gly Asn Trp Asp Thr Ser Lys Ala Ile Gly Pro Met Phe Asn
610 615 620

Gln Val Val Tyr Gln Tyr Pro Thr Trp Tyr Tyr Asp Val Ser Val Pro
625 630 635 640

Ala Gly Thr Thr Ile Gln Phe Lys Phe Ile Lys Lys Asn Gly Asn Thr
645 650 655

Ile Thr Trp Glu Gly Gly Ser Asn His Thr Tyr Thr Val Pro Ser Ser
660 665 670

Ser Thr Gly Thr Val Ile Val Asn Trp Gln Gln
675 680

<210> 3
<211> 683
<212> PRT
<213> Thermoanaerobacter sp.

<400> 3

Ala Pro Asp Thr Ser Val Ser Asn Val Val Asn Tyr Ser Thr Asp Val
1 5 10 15

Ile Tyr Gln Ile Val Thr Asp Arg Phe Leu Asp Gly Asn Pro Ser Asn
20 25 30

Asn Pro Thr Gly Asp Leu Tyr Asp Pro Thr His Thr Ser Leu Lys Lys
35 40 45

Tyr Phe Gly Gly Asp Trp Gln Gly Ile Ile Asn Lys Ile Asn Asp Gly
50 55 60

Tyr Leu Thr Gly Met Gly Ile Thr Ala Ile Trp Ile Ser Gln Pro Val
65 70 75 80

Glu Asn Ile Tyr Ala Val Leu Pro Asp Ser Thr Phe Gly Gly Ser Thr
85 90 95

Ser Tyr His Gly Tyr Trp Ala Arg Asp Phe Lys Lys Thr Asn Pro Phe
100 105 110

Phe Gly Ser Phe Thr Asp Phe Gln Asn Leu Ile Ala Thr Ala His Ala
115 120 125

```
His Asn Ile Lys Val Ile Ile Asp Phe Ala Pro Asn His Thr Ser Pro
    130                 135                 140

Ala Ser Glu Thr Asp Pro Thr Tyr Gly Glu Asn Gly Arg Leu Tyr Asp
145                 150                 155                 160

Asn Gly Val Leu Leu Gly Gly Tyr Thr Asn Asp Thr Asn Gly Tyr Phe
                165                 170                 175

His His Tyr Gly Gly Thr Asn Phe Ser Ser Tyr Glu Asp Gly Ile Tyr
            180                 185                 190

Arg Asn Leu Phe Asp Leu Ala Asp Leu Asp Gln Gln Asn Ser Thr Ile
        195                 200                 205

Asp Ser Tyr Leu Lys Ala Ala Ile Lys Leu Trp Leu Asp Met Gly Ile
    210                 215                 220

Asp Gly Ile Arg Met Asp Ala Val Lys His Met Ala Phe Gly Trp Gln
225                 230                 235                 240

Lys Asn Phe Met Asp Ser Ile Leu Ser Tyr Arg Pro Val Phe Thr Phe
                245                 250                 255

Gly Glu Trp Tyr Leu Gly Thr Asn Glu Val Asp Pro Asn Asn Thr Tyr
            260                 265                 270

Phe Ala Asn Glu Ser Gly Met Ser Leu Leu Asp Phe Arg Phe Ala Gln
        275                 280                 285

Lys Val Arg Gln Val Phe Arg Asp Asn Thr Asp Thr Met Tyr Gly Leu
    290                 295                 300

Asp Ser Met Ile Gln Ser Thr Ala Ala Asp Tyr Asn Phe Ile Asn Asp
305                 310                 315                 320

Met Val Thr Phe Ile Asp Asn His Asp Met Asp Arg Phe Tyr Thr Gly
                325                 330                 335

Gly Ser Thr Arg Pro Val Glu Gln Ala Leu Ala Phe Thr Leu Thr Ser
            340                 345                 350

Arg Gly Val Pro Ala Ile Tyr Tyr Gly Thr Glu Gln Tyr Met Thr Gly
        355                 360                 365

Asn Gly Asp Pro Tyr Asn Arg Ala Met Met Thr Ser Phe Asp Thr Thr
    370                 375                 380

Thr Thr Ala Tyr Asn Val Ile Lys Lys Leu Ala Pro Leu Arg Lys Ser
385                 390                 395                 400
```

Asn Pro Ala Ile Ala Tyr Gly Thr Gln Lys Gln Arg Trp Ile Asn Asn
405 410 415

Asp Val Tyr Ile Tyr Glu Arg Gln Phe Gly Asn Asn Val Ala Leu Val
420 425 430

Ala Ile Asn Arg Asn Leu Ser Thr Ser Tyr Tyr Ile Thr Gly Leu Tyr
435 440 445

Thr Ala Leu Pro Ala Gly Thr Tyr Ser Asp Met Leu Gly Gly Leu Leu
450 455 460

Asn Gly Ser Ser Ile Thr Val Ser Ser Asn Gly Ser Val Thr Pro Phe
465 470 475 480

Thr Leu Ala Pro Gly Glu Val Ala Val Trp Gln Tyr Val Ser Thr Thr
485 490 495

Asn Pro Pro Leu Ile Gly His Val Gly Pro Thr Met Thr Lys Ala Gly
500 505 510

Gln Thr Ile Thr Ile Asp Gly Arg Gly Phe Gly Thr Thr Ala Gly Gln
515 520 525

Val Leu Phe Gly Thr Thr Pro Ala Thr Ile Val Ser Trp Glu Asp Thr
530 535 540

Glu Val Lys Val Lys Val Pro Ala Leu Thr Pro Gly Lys Tyr Asn Ile
545 550 555 560

Thr Leu Lys Thr Ala Ser Gly Val Thr Ser Asn Ser Tyr Asn Asn Ile
565 570 575

Asn Val Leu Thr Gly Asn Gln Val Cys Val Arg Phe Val Val Asn Asn
580 585 590

Ala Thr Thr Val Trp Gly Glu Asn Val Tyr Leu Thr Gly Asn Val Ala
595 600 605

Glu Leu Gly Asn Trp Asp Thr Ser Lys Ala Ile Gly Pro Met Phe Asn
610 615 620

Gln Val Val Tyr Gln Tyr Pro Thr Trp Tyr Tyr Asp Val Ser Val Pro
625 630 635 640

Ala Gly Thr Thr Ile Glu Phe Lys Phe Ile Lys Lys Asn Gly Ser Thr
645 650 655

Val Thr Trp Glu Gly Gly Tyr Asn His Val Tyr Thr Thr Pro Thr Ser
660 665 670

Gly Thr Ala Thr Val Ile Val Asp Trp Gln Pro
675 680